Optimal Prediction of Stiff Oscillatory Systems

Anton P. Kast

Applied Numerical Algorithms Group
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

May 5, 1999

Abstract

We consider some large systems of differential equations that have been introduced as model many-body problems. These systems have solutions that oscillate on a wide range of time scales. We apply the formalism of optimal prediction to these systems, using conditional expectations of the equations of motion to construct effective equations for the most slowly-varying quantities. We verify the accuracy of the effective equations in examples, comparing solutions of the original and new systems, and we show that the new equations give accurate answers for slow variables with relatively little computational effort.

This work was supported in part by the Department of Energy's Office of Advanced Computing Research and its Mathematical, Information, and Computational Sciences Division, under Contract No. DE-AC03-76SF00098.

In a recent paper[1], Stuart and Warren considered a particular Hamiltonian dynamical system as a model of a particle interacting with a heat bath. This dynamical system consisted of many particles connected by springs, and by choosing the masses of the particles to vary over a wide range, these authors caused stiffness in Hamilton's equations of motion that mimicked the stiffness which limits more realistic molecular dynamics computations. They then introduced a variety of numerical schemes for integrating the equations of motion, and asked what would happen if the schemes were grossly underresolved in time. Stuart and Warren showed that schemes could be found which yielded correct answers for slowly-varying degrees of freedom, even when most of the dynamics was underresolved in time (i.e., even when the time step was much larger than the periods of most normal modes of oscillation).

This observation, that a scheme may be optimized to work well at poor resolution, is similar to the claims of optimal prediction [3, 4]. Optimal prediction is a formalism for reducing a large system of differential equations into a smaller system of differential equations. The smaller system is designed to yield expectations of solutions to the larger system and to be computationally practical when the larger system is not. Since Stuart and Warren

have found schemes for large, stiff systems that work with big time steps, it is natural to ask whether there are smaller systems of differential equations (just describing the slower modes) that would work at these big time steps.

In this paper, we apply optimal prediction to the Stuart-Warren model. We reduce their original, large system of differential equations to a much smaller system, and show that the resolved integration of the smaller system reproduces the benefits of their large-Δt schemes. We then add more interactions to the original mass-and-spring model, so that every mass interacts with every other mass. The same optimal prediction formalism applies to this more elaborate example, and defines a renormalization of spring constants. We demonstrate the computational savings of the reduced equations for this new model as well.

Our methods are not limited to the model problems. Our results are potentially relevant for all mechanics problems involving oscillatory motion and a wide range of time scales. For a review of problems and methods of this type, including molecular dynamics, orbital mechanics, and electronic oscillators, see [9].

1 Formulas for conditional expectations

In this paper, we assume that some dynamical variables are of interest and others are irrelevant and unknown. We take all the unknown quantities to be random variables, and we assume we are given a prior probability distribution to describe their statistics. Since we are considering Hamiltonian equations, we take this prior distribution to be the canonical ensemble.

The canonical ensemble is defined by the probability density, $P = e^{-H}$ where H is the Hamiltonian. If H is quadratic in canonical variables, as in the examples presented here, then P is a Gaussian distribution.

Our general approach is to pick some particular variables to compute, which we call "collective variables," and to treat the values of these variables as conditions when we compute expectations of everything else. Since P will be Gaussian, we will be interested in conditional expectations with respect to Gaussian distributions. This section presents general formulas for computing such expectations.

Let $x_1 \ldots x_N$ be Gaussian random variables distributed with density

$$P(x_1, \ldots, x_N) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i A_{ij} x_j - \frac{1}{2} \sum_{i=1}^{N} b_i x_i \right). \qquad (1)$$

We denote expectations with respect to this density by $\langle \cdot \rangle$, and $\langle x_i \rangle =$

$\sum_{j=1}^{N} A_{ij}^{-1} b_j$. Now suppose that $x_1 \ldots x_n$ are given for all $n < N$. The conditional expectations of $x_{n+1} \ldots x_N$ conditioned by $x_1 \ldots x_n$ are denoted $\langle x_i \rangle_n$, $i = n+1, \ldots, N$ and are given explicitly by

$$\langle x_i \rangle_n = \langle x_i \rangle + \sum_{\mu=1}^{n} \sum_{\nu=1}^{n} A_{i\mu}^{-1} M_{\mu\nu}^{-1} (x_\nu - \langle x_\nu \rangle), \qquad i = n+1, \ldots, N \tag{2}$$

where $M_{\mu\nu} = A_{\mu\nu}^{-1}$ for $\mu, \nu = 1, \ldots, n$ and M^{-1} is the inverse of the $n \times n$ (not $N \times N$) matrix (see [3, 4] for details).

In this paper, all matrices A_{ij} will be of the form,

$$A_{ij} = a\delta_{ij} + b \tag{3}$$

for some real numbers $a \neq 0$ and $b \neq -a/N$. These matrices are always invertible (they form a group) and the formula for the inverse is

$$A_{ij}^{-1} = \frac{1}{a}\delta_{ij} - \frac{b/a}{a + Nb}. \tag{4}$$

If these x_i variables are functions of time, and if they obey a system of N first-order differential equations

$$\dot{x}_i = f_i(x_1, \ldots, x_N), \qquad i = 1, \ldots, N \tag{5}$$

then the first approximation of optimal prediction states that if $x_1 \ldots x_n$ are known at an initial time and $x_{n+1} \ldots x_N$ are not, then

$$\dot{x}_\mu = \langle f_\mu(x_1, \ldots, x_N) \rangle_n, \qquad \mu = 1, \ldots, n \tag{6}$$

is an accurate system of $n < N$ differential equations for the $x_1 \ldots x_n$ on average, at least for short times. Note that the right-hand sides of (6) are functions of $x_1 \ldots x_n$ through conditions on the expectation, and not of $x_{n+1} \ldots x_N$.

2 The original model

Stuart and Warren [1] (see also [2] and [8]) considered a one-dimensional collection of particles connected by springs. There was one distinguished particle, with mass 1, coordinate Q and momentum P. The distinguished particle was connected, by springs of spring constant k, to N other particles with masses k/j^2, coordinates q_j and momenta p_j, $j = 1 \ldots N$ representing a heat bath (see Ford and Kac [6]).

The motion of this collection of particles and springs is defined by the Hamiltonian

$$H(Q, P; q_1, \ldots, q_N; p_1, \ldots, p_N) = \frac{1}{2}(Q^2 + P^2) + \sum_{j=1}^{N} \left[\frac{p_j^2}{2m_j} + \frac{1}{2}k(Q - q_j)^2 \right] \quad (7)$$

where (Q, P) and (q_j, p_j) are canonically conjugate dynamical variables for

$j = 1, \ldots, N$ and $m_j = k/j^2$. The associated equations of motion are

$$\begin{aligned} \dot{Q} &= P \\ \dot{P} &= -Q + k\sum_{j=1}^{N}(q_j - Q) \\ \dot{q}_j &= p_j/m_j, \qquad\qquad j = 1, \ldots, N \\ \dot{p}_j &= k(Q - q_j), \qquad\qquad j = 1, \ldots, N \end{aligned} \tag{8}$$

If Q were fixed, then each q_j would oscillate harmonically with frequency $\omega_j = j$. A discretization of the $2N + 2$ equations (8) would therefore be resolved in time if $\Delta t \ll \frac{2\pi}{\omega_N} = \frac{2\pi}{N}$. If this condition on Δt were violated, then the result of the computation would depend on how the equations were discretized. The intriguing result of [1] is that some schemes will give the right evolution for Q and P when $\Delta t \gtrsim \frac{2\pi}{N}$ and others will not. For instance, if the scheme is

$$\begin{aligned} \frac{Q^{n+1} - Q^n}{\Delta t} &= P^{n+1} \\ \frac{P^{n+1} - P^n}{\Delta t} &= -Q^n + k\sum_{j=1}^{N}(q_j^{n+\sigma} - Q^n) \\ \frac{q_j^{n+1} - q_j^n}{\Delta t} &= p_j^{n+1}/m_j \\ \frac{p_j^{n+1} - p_j^n}{\Delta t} &= k(Q^n - q_j^n) \end{aligned} \tag{9}$$

then $\sigma = 0$ (a symplectic method) gives the right answer for Q and P, but $\sigma = 1$ (another convergent method) does not.

Figure 1 shows a fully-resolved calculation ($\Delta t = 10^{-2}/N$) of $P(t)$ starting

from $P(0) = 0$, $Q(0) = 1.5$, with $q_j(0)$ and $p_j(0)$ chosen randomly from the canonical ensemble, i.e., chosen with probability density e^{-H}. This figure also shows a calculation made from the same initial condition, but using the symplectic scheme (equation (9) with $\sigma = 0$) and $\Delta t = 1/N$, a time step one hundred times larger. This Δt is too large to resolve most normal modes, but the two calculations overlap; they are indistinguishable. Clearly, the fastest oscillations in the model (8) do not need to be resolved in order to get $P(t)$ right.

3 Optimal prediction of the original model

If we choose our collective variables to be Q, P, $q_1, \ldots, q_n$ and $p_1, \ldots, p_n$ for some $0 \le n \le N$, then the optimal prediction equations of the model (8) have the same form as the original equations. Taking the conditional expectations of the right-hand sides of (8) and evaluating the expectations using (2), we find that

$$\begin{aligned} \dot{Q} &= P \\ \dot{P} &= -Q + k \sum_{\mu=1}^{n} (q_\mu - Q) \\ \dot{q}_\mu &= p_\mu / m_\mu, \qquad\qquad \mu = 1, \ldots, n \\ \dot{p}_\mu &= k(Q - q_\mu), \qquad\qquad \mu = 1, \ldots, n \end{aligned} \tag{10}$$

on average. It comes as no surprise, therefore, that the motion of Q can be computed with large Δt: pick the Δt desired, find an $n \ll N$ such that $\Delta t \ll \frac{2\pi}{n}$, and perform a resolved integration of (10) with this n and Δt. A reasonable approximation for $P(t)$ is guaranteed at least for short times.

Figure 2 duplicates the fully-resolved solution from Figure 1, and also shows the solution to the same problem as computed by a resolved integration of (10), instead of an underresolved implementation of (9). Both calculations used the same large $\Delta t = 10^{-2}/n = 1/N$. The underresolved calculation duplicates the exact solution in high-frequency detail, but the optimal prediction involves less computation since it is n-dimensional instead of N-dimensional. The optimal prediction has the additional advantage that it did not use the initial data $q_{n+1}(0) \ldots q_N(0)$, $p_{n+1}(0) \ldots p_N(0)$ and may claim to be an average answer over all possible values of these data.

4 A new model

Realistic molecular dynamics involves more complex interactions than are present in the model (8). In particular, in reality every particle would interact with every other, and the interactions would be nonlinear.

Nonlinear interactions require perturbative treatment in optimal predic-

tion. There has been recent progress on this problem (see [5]), but nonlocal interaction is simple to analyze. We therefore consider the generalization of the model (8) where every $q_1 \dots q_N$ is coupled to every other $q_1 \dots q_N$ by a spring:

$$\begin{aligned} H(Q,P;q_1,\dots,q_N;p_1,\dots,p_N) \\ &= \frac{1}{2}(Q^2+P^2) + \sum_{j=1}^{N}\left[\frac{p_j^2}{2m_j} + \frac{1}{2}k_Q(Q-q_j)^2\right] \\ &\qquad + \frac{1}{2}k_q\sum_{j=1}^{N}\sum_{l=j+1}^{N}(q_j-q_l)^2 \end{aligned} \tag{11}$$

$$\begin{aligned} \dot{Q} &= P \\ \dot{P} &= -Q + k_Q\sum_{j=1}^{N}(q_j - Q) \\ \dot{q}_j &= p_j/m_j, \qquad\qquad j=1,\dots,N \\ \dot{p}_j &= k_Q(Q-q_j) + k_q\sum_{l=1}^{N}(q_l-q_j), \quad j=1,\dots,N. \end{aligned} \tag{12}$$

We have introduced two new spring constants in this formula: k_Q, for the couplings between the distinguished particle and all the others; and k_q, for all the couplings among the others.

5 Optimal prediction of the new model

We derive the optimal prediction equations of the system (12) for Q, P, $q_1 \dots q_n$, $p_1 \dots p_n$ by averaging over $q_{n+1} \dots q_N$, $p_{n+1} \dots p_N$. When Q and P

are constrained, the canonical probability density e^{-H} remains Gaussian, and it factorizes in the q's and p's:

$$P_{QP}(q_1, \ldots, q_N; p_1, \ldots, p_N) \propto \exp\left(-\sum_{j=1}^{N} \frac{p_j^2}{m_j}\right) \exp\left(-\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}(q_j - Q)A_{jk}(q_k - Q)\right) \quad (13)$$

where

$$A_{jk} = (k_Q + Nk_q)\delta_{jk} - k_q. \quad (14)$$

If, for some $n < N$, $q_1 \ldots q_n$ and $p_1 \ldots p_n$ are also constrained, then the conditional expectations of $q_{n+1} \ldots q_N$ and $p_{n+1} \ldots p_N$ are given according to (2) by

$$\langle p_j \rangle_n = \langle p_j \rangle = 0 \quad (15)$$

$$\langle q_j \rangle_n = \langle q_j \rangle + \sum_{\mu=1}^{n}\sum_{\nu=1}^{n} A_{j\mu}^{-1} M_{\mu\nu}^{-1}(q_\nu - \langle q_\nu \rangle) \quad (16)$$

where $M_{\mu\nu} = A_{\mu\nu}^{-1}$, $\mu, \nu = 1, \ldots, n$ are the components of the $n \times n$ matrix M.

Taking conditional expectations of the equations of motion (12), these

elementary expectations imply that

$$\begin{aligned}
\langle \dot{Q} \rangle_n &= P \\
\langle \dot{P} \rangle_n &= -Q + k'_Q \sum_{\mu=1}^{n} (q_\mu - Q) \\
\langle \dot{q}_\mu \rangle_n &= p_\mu / m_\mu, \qquad\qquad \mu = 1, \ldots, n \\
\langle \dot{p}_\mu \rangle_n &= k'_Q (Q - q_\mu) + k'_q \sum_{\nu=1}^{n} (q_\nu - q_\mu), \qquad \mu = 1, \ldots, n
\end{aligned} \tag{17}$$

where

$$k'_q = C k_q, \qquad k'_Q = C k_q, \tag{18}$$

and

$$C = 1 + (N - n) \frac{k_q}{k_Q + n k_q}. \tag{19}$$

The first approximation of optimal prediction (6) yields the system

$$\begin{aligned}
\dot{Q} &= P \\
\dot{P} &= -Q + k'_Q \sum_{\mu=1}^{n} (q_\mu - Q) \\
\dot{q}_\mu &= p_\mu / m_\mu, \qquad\qquad \mu = 1, \ldots, n \\
\dot{p}_\mu &= k'_Q (Q - q_\mu) + k'_q \sum_{\nu=1}^{n} (q_\nu - q_\mu), \qquad \mu = 1, \ldots, n
\end{aligned} \tag{20}$$

which is interpreted as an approximate evolution rule for averages of Q, P, $q_1 \ldots q_n$ and $p_1 \ldots p_n$.

The only difference between the reduced equations (20) and the original equations (12) is in the coupling constants, k_Q and k_q. This is unexpected.

The reduced equations can be derived by ignoring the lighter particles and just rescaling the spring constants. It amounts to a dynamical analogue of renormalization in physics [7].

We performed a more rigorous test of the new model, comparing it to an actual mean evolution. The results are shown in Figure 3. We once again picked initial data $Q(0) = 1.5$, $P(0) = 0$, and chose $q_1 \dots q_n$, $p_1 \dots p_n$ ($n = 10$) from the canonical distribution for N particles ($N = 1000$ at $k_Q = k_q = 1$. We then generated an ensemble of 100 sets of values for $q_{n+1} \dots q_N$, $p_{n+1} \dots p_N$, and for each set integrated the equations (12). Averaging over all 100 solutions yielded the first curve for $P(t)$. We then discarded the ensemble and used the original $q_1 \dots q_n$, $p_1 \dots p_n$ as initial conditions for the reduced system (20), which we integrated with $\Delta t = 10^{-2}/n = 1/N$. This Δt is small enough to resolve the reduced dynamics but much too large to resolve the original dynamics. The solution for $P(t)$ from (20) is the second curve. Finally, for comparison we performed the naive experiment of simply truncating the big system (12) to n degrees of freedom, effectively ignoring the lighter particles without rescaling k_Q and k_q by the factor C, which in this case is 9.1818... This produced the third curve.

The figure shows that the reduced system accurately predicts the average

evolution of $P(t)$, and it does so at a computational savings of 99 percent of the degrees of freedom and a factor of roughly 100 in time step.

6 Conclusions

We have shown how some large systems of differential equations with many time scales may be systematically reduced to smaller systems with weaker time step conditions by the probabilistic technique of optimal prediction.

The similarities between our examples and practical models of molecular dynamics suggest that our methods might be useful for more realistic systems of equations.

7 Acknowledgements

The author thanks...

References

[1] A.M. Stuart and J.O. Warren. Analysis and experiments for a computational model of a heat bath. SCCM Technical Report SCCM-98-17 (1998). Available from http://www-sccm.stanford.edu.

[2] B. Cano, A.M. Stuart, E. Süli and J.O. Warren. Stiff oscillatory systems, delta jumps and white noise. SCCM Technical Report SCCM-99-01 (1999). Available from http://www-sccm.stanford.edu.

[3] A.J. Chorin, A. Kast and R. Kupferman. Unresolved computation and optimal predictions. *Comm. Pure Appl. Math.*, in press.

[4] A.J. Chorin, A. Kast and R. Kupferman. On the prediction of large-scale dynamics using underresolved computations. *AMS Contemporary Mathematics*, in press.

[5] A.J. Chorin and D. Levy. In preparation.

[6] G.W. Ford and M. Kac. On the quantum Langevin equation. *J. Stat. Phys.* **46**, 803 (1987).

[7] J.C. Collins. Renormalization: an introduction to renormalization, the renormalization group, and the operator-product expansion. New York: Cambridge University Press (1984).

[8] G.W. Ford, J.T. Lewis and R.F. O'Connell. Quantum Langevin equation. *Phys. Rev. A.* **37**, 4419 (1988).

[9] L.R. Petzold, L.O. Jay and J. Yen. Numerical solution of highly oscillatory differential equations. *Acta Numerica* **6**, 437 (1997).

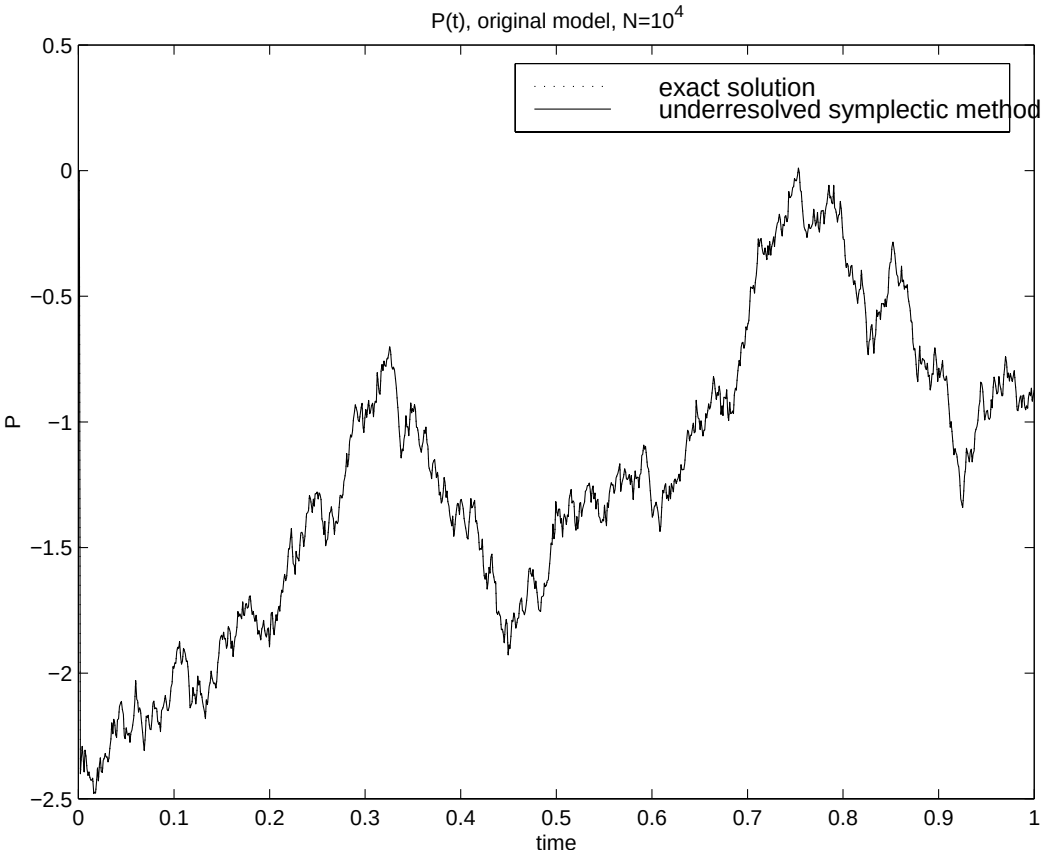

Figure 1: The evolution of $P(t)$ determined in two ways: by solving the equations of motion (8) with $\Delta t = 10^{-2}/N$ (exact solution); and by applying the symplectic scheme (9) with $\Delta t = 1/N$ (underresolved symplectic method). For these calculations, $N = 10^4$ and $k_Q = k_q = 1$. The two solutions overlap on the scale of this plot.

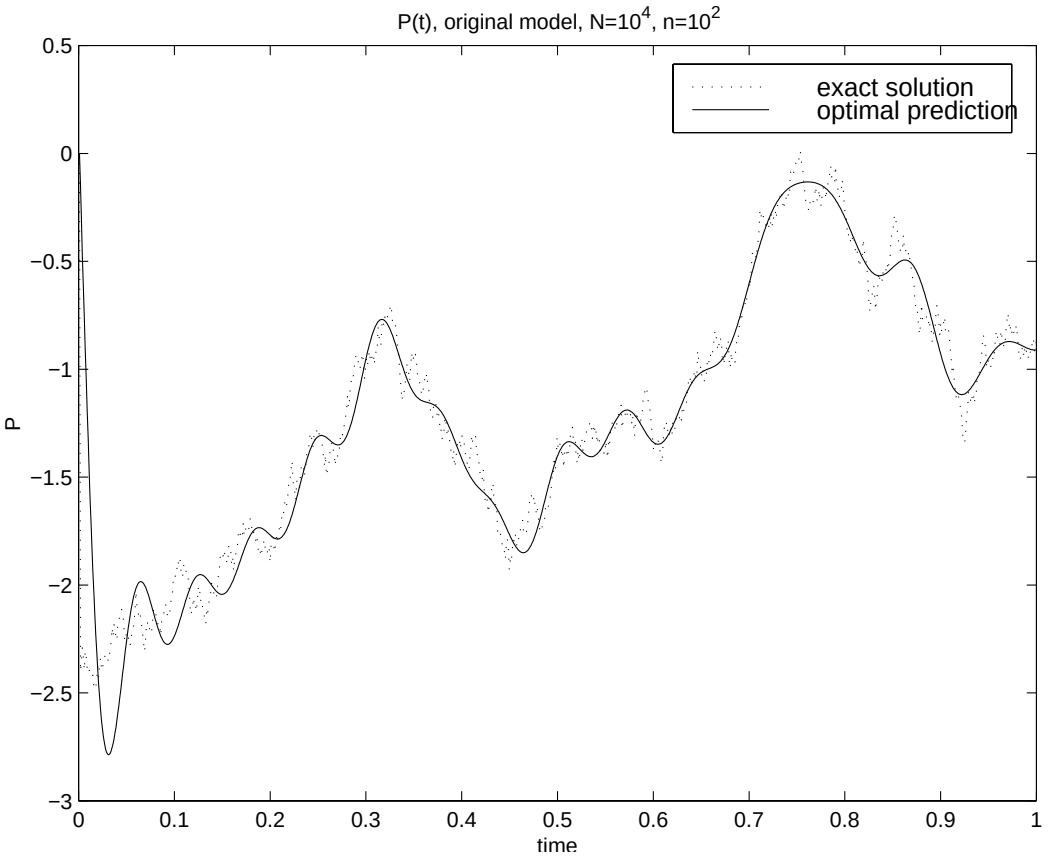

Figure 2: The evolution of P(t) determined in two ways: by solving the equations of motion (8) for $N = 10^4$ with $\Delta t = 10^{-2}/N$ (exact evolution); and by solving the reduced equations (20) with $n = 10^2$ and $\Delta t = 1/N = 10^{-2}/n$ (optimal prediction). For these calculations, $k_Q = k_q = 1$.

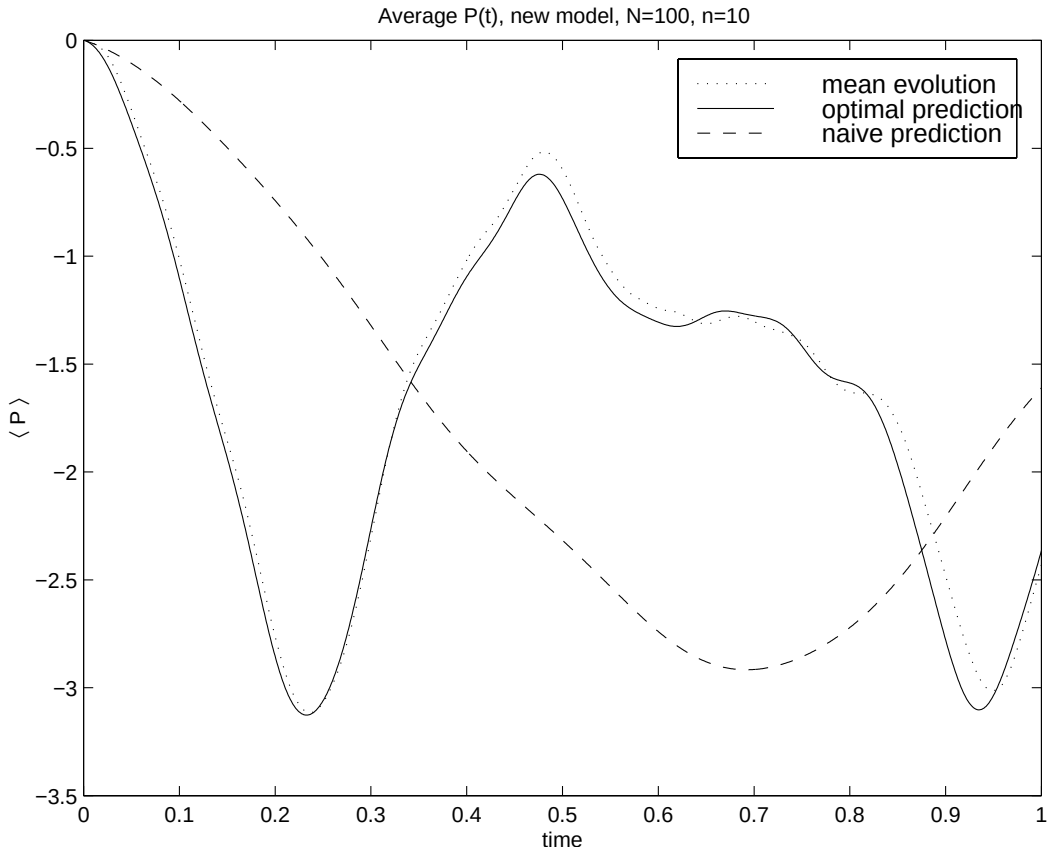

Figure 3: The *average* evolution of P(t) determined in three ways: by solving the equations of motion (12) for 100 different initial conditions, with $N = 10^3$, $\Delta t = 10^{-2}/N$, and then averaging all 100 solutions (mean evolution); by solving the reduced equations (20) once, with $n = 10$ and $\Delta t = 1/N = 10^{-2}/n$ (optimal prediction); and by solving the reduced equations (20) once with $n = 10$, $\Delta t = 10^{-2}/n$ but without rescaling the couplings, k_Q and k_q, by the factor $C = 9.1818...$